



Abstract

[042] A text-to-speech synthesizer employs database that includes units. For each unit there is a collection of unit selection parameters and a plurality of frames. Each frame has a set of model parameters derived from a base speech frame, and a speech frame synthesized from the frame's model parameters. A text to be synthesized is converted to a sequence of desired unit features sets, and for each such set the database is perused to retrieve a best-matching unit. An assessment is made whether modifications to the frames are needed, because of discontinuities in the model parameters at unit boundaries, or because of differences between the desired and selected unit features. When modifications are necessary, the model parameters of frames that need to be altered are modified, and new frames are synthesized from the modified model parameters and concatenated to the output. Otherwise, the speech frames previously stored in the database are retrieved and concatenated to the output.